

What is claimed is:

1. A method for maintaining records in a database comprising:

(a) receiving at least a collection of first data items and a collection of second data items, at least one of the first data items being associated with one of the second data items;

(b) disposing the first data items in a plurality of fields arranged in a predetermined format to form a first assemblage;

(c) disposing the second data items in a plurality of fields arranged in the predetermined format to form a second assemblage;

(d) modifying selected ones of the first data items and the second data items to conform to predetermined nomenclature;

(e) maintaining a record, in the database, having a plurality of fields arranged in the predetermined format; and

(f) including in a field in the record a selected one of (1) the first data items in the predetermined nomenclature and (2) the second data items in the predetermined nomenclature.

2. The method of Claim 1 wherein at (b) and (c) the first data items and the second data items are separated, rearranged, or combined to form the first assemblage and the second assemblage respectively.

3. The method of Claim 1 wherein at (f) the first data items and the second data items are assigned accuracy rankings and are selected based upon their accuracy rankings.

4. A system for maintaining records in a database comprising:

a communications interface for receiving at least a collection of first data items

and a collection of second data items, at least one of the first data items being associated with one of the second data items;

a converter for disposing the first data items in a plurality of fields arranged in a predetermined format to form a first assemblage, and for disposing the second data  
5 items in a plurality of fields arranged in the predetermined format to form a second assemblage;

a normalizer for modifying selected ones of the first data items and the second data items to conform to predetermined nomenclature;

a database for maintaining a record, in the database, having a plurality of fields  
10 arranged in the predetermined format; and

a data masher for including in a field in the record a selected one of (1) the first data items in the predetermined nomenclature and (2) the second data items in the predetermined nomenclature.

15 5. The system of Claim 4 wherein each of the of the first assemblage of data items and the second assemblage of data items in the predetermined nomenclature are associated with data accuracy rankings.

20 6. The system of Claim 5 wherein the data masher selected the first data items and the second data items based upon the data accuracy rankings.

7. A system for maintaining records in a database comprising:

a converter for converting a collection of first data items and a collection of second data items to a plurality of fields arranged in a predetermined format to form a  
25 first assemblage of data items and a second assemblage of data items, at least one of the first data items being associated with one of the second data items;

a normalizer for converting selected ones of the first assemblage of data items

and the second assemblage of data items to conform to a predetermined nomenclature;

a database for maintaining a record, in the database, having a plurality of fields arranged in the predetermined format; and

a data masher for selecting at least one data item from the first assemblage of data items in the predetermined nomenclature, selecting data items from the second assemblage of data items in the predetermined nomenclature, and combining the at least one data items and the data items from the second assemblage to create the record.

8. The system of Claim 7 wherein the converter separates, rearranges, or combines data to form the first assemblage of data items and the second assemblage of data items.

9. The system of Claim 7 wherein each of the of the first assemblage of data items and the second assemblage of data items in the predetermined nomenclature are associated with data accuracy rankings.

10. The system of Claim 9 wherein the data masher selected the first data items and the second data items based upon the data accuracy rankings.

11. A system for maintaining records in a database comprising:  
a converter for converting a collection of first records having fields and data items into uniform records having fields and data items organized into a uniform format and where each of the uniform records have one or more of the data items conforming to a predetermined nomenclature and a balance of the one or more data items not conforming to the predetermined nomenclature;

a normalizer for converting the balance of the one or more data items to the

predetermined nomenclature, producing a collection of second records each having the one or more of the data items that conform to a predetermined nomenclature and the balance of the one or more data items that are in the predetermined nomenclature; and

a data masher for selecting equivalent records from the collection of second  
 5 records and selecting data items from the equivalent records for combining into a third record having the selected data items arranged in the uniform format and stored in the database.

12. The system of Claim 11 wherein the converter separates, rearranges, or  
 10 combines the fields and the data items, from the first records, into the uniform format.

13. The system of Claim 11 wherein before the normalizer converts the balance of the one or more data items it separates the one or more data items into components.

14. The system of Claim 11 wherein the data masher selects equivalent records  
 15 by comparing the data items in the second records on a field-by-field basis and determines a confidence level value of the second records based upon the similarities of the data items in the second records, and determining that the second records are equivalent if the confidence level values are above a predetermined threshold level  
 20 value.

15. Software recordable in a tangible medium which includes machine readable instructions for performing a process for building a database, which stores records corresponding to a plurality of data items, the process comprising:

25 (a) receiving at least a collection of first data items and a collection of second data items, at least one of the first data items being associated with one of the second data items;

(b) disposing the first data items in a plurality of fields arranged in a predetermined format to form a first assemblage;

(c) disposing the second data items in a plurality of fields arranged in the predetermined format to form a second assemblage;

5 (d) modifying selected ones of the first data items and the second data items to conform to predetermined nomenclature;

(e) maintaining a record, in the database, having a plurality of fields arranged in the predetermined format; and

10 (f) including in a field in the record a selected one of (1) the first data items in the predetermined nomenclature and (2) the second data items in the predetermined nomenclature.

15 16. The process of Claim 15 wherein at (b) and (c) the first data items and the second data items are separated, rearranged, or combined to form the first assemblage and the second assemblage respectively.

20 17. The process of Claim 15 wherein at (f) the first data items and the second data items are assigned accuracy rankings and are selected based upon their accuracy rankings.

25 18. A method for maintaining records in a database comprising:  
converting a collection of first records having fields and data items into uniform records having fields and data items and where each of the uniform records have one or more of the data items conforming to a predetermined nomenclature and a balance of the one or more data items not conforming to the predetermined nomenclature;

converting the balance of the one or more data items to the predetermined nomenclature, producing a collection of second records each having the one or more of

the data items that conform to a predetermined nomenclature and the balance of the one or more data items that are in the predetermined nomenclature; and

selecting equivalent records from the collection of second records and selecting the data items from the equivalent records for combining into a third record having the selected data items and stored in the database.

19. The method of Claim 18 wherein the fields and the data items, from the first records, are separated, rearranged, or combined.

20. The method of Claim 18 wherein the balance of the one or more data items are separated into components and converted into the predetermined nomenclature.

21. The method of Claim 18 wherein equivalent records are selected by comparing the data items in the second records on a field-by-field basis and determining a confidence level value of the second records based upon the similarities of the data items in the second records, and determining that the second records are equivalent if the confidence level values are above a predetermined threshold level value.